Roc, a Ras/GTPase domain in complex proteins

Bosgraaf, Leonard; Haastert, Peter J.M. van

Published in:
Biochimica et Biophysica Acta (BBA) - Molecular Cell Research

DOI:
10.1016/j.bbamcr.2003.08.008

Publication date:
2003

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochimica et Biophysica Acta 1643 (2003) 5–10
BIOCHIMICA ET BIOPHYSICA ACTA
BBA
www.bba-direct.com

Short sequence-paper

Roc, a Ras/GTPase domain in complex proteins

Leonard Bosgraaf, Peter J.M. Van Haastert*

Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received 13 May 2003; received in revised form 12 August 2003; accepted 27 August 2003

Abstract

We identified a novel group of the Ras/GTPase superfamily, termed Roc, that is present as domain in complex proteins together with other domains, including leucine-rich repeats (LRRs), ankyrin repeats, WD40 repeats, kinase domains, RasGEF and RhoGAP domains. Roc is always succeeded by a novel 300–400-amino-acid-long domain, termed COR. Proteins with Roc/COR are present in prokaryotes, *Dictyostelium*, plants and metazoa.

Keywords: Roc; Ras/GTPase; Protein

Members of the Ras/GTPase superfamily are small proteins that usually do not possess additional domains [1,2]. We recently characterised a complex protein in the slime mould *Dictyostelium discoideum* that contains Ras/GTPase, MAPKKkinase and several other domains [3]. This protein was termed GbpC for c*G*MP *b*inding *p*rotein C. Knock-out studies indicate that this protein is essential for efficient chemotaxis, cell polarization and myosin II phosphorylation [4]. The human protein KIAA1790 was also found, which contains homology to the kinase and Ras/GTPase domains, and also shows substantial homology to the 45-kDa region between these domains. Here we present the identification of several other proteins that have a similar domain architecture. As discussed below, the Ras domains in these proteins form a novel subfamily of small GTPases. We termed this domain Roc for *R*as *o*f *c*omplex proteins. Furthermore, we propose the name COR (*C*-terminal *o*f *Roc*) for the region between the Roc and the kinase domains, and Roco protein for the family of proteins that contain both Roc and COR domains.

1. Searching for Roco proteins

The amino acid sequences of the Roc/COR domain of *Dictyostelium* GbpC (amino acid 331–877) and the human homologue KIAA1790 (amino acid 1–585) were used to search the SWISS-PROT (http://www.ncbi.nlm.nih.gov/BLAST/) and *Dictyostelium* (http://www.sanger.ac.uk/Projects/D_discoideum/ and http://dicty.sdsc.edu/) sequence databases using BLASTP and TBLASTN. This revealed a number of homologous sequences, which were used for further BLAST and PSI-BLAST searches. The identified 40 independent sequences were submitted to the online MEME/MAST system (http://meme.sdsc.edu/meme/website/) [5] for searching still other potential weak scoring proteins. MEME/MAST found all previous hits from the Genbank/SWISS-PROT database, but no additional sequences with significant identity to the Roc/COR domains or the COR domain alone.

Roco proteins were found in prokaryotes, *Dictyostelium*, plants, and metazoa but not in *Plasmodium* and yeast. Remarkably, we found that the Roc and COR domains are always connected and we did not observe any proteins containing either the Roc or the COR domain alone, indicating that these two domains might function as one inseparable unit.

2. Domain architecture of Roco proteins

The identified Roco proteins fall apart in three groups (see Fig. 1). The first group was found in mammals (MASL1), the plant *Arabidopsis* and in some prokaryotes, including several archea and cyanobacteria. In these proteins the Roc domain is always preceded by 8–18 leucine-

* Corresponding author. Tel.: +31-503634172; fax: +31-503634165.
E-mail address: P.J.M.van.Haastert@chem.rug.nl (P.J.M. Van Haastert).

0167-4889/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.bbamcr.2003.08.008

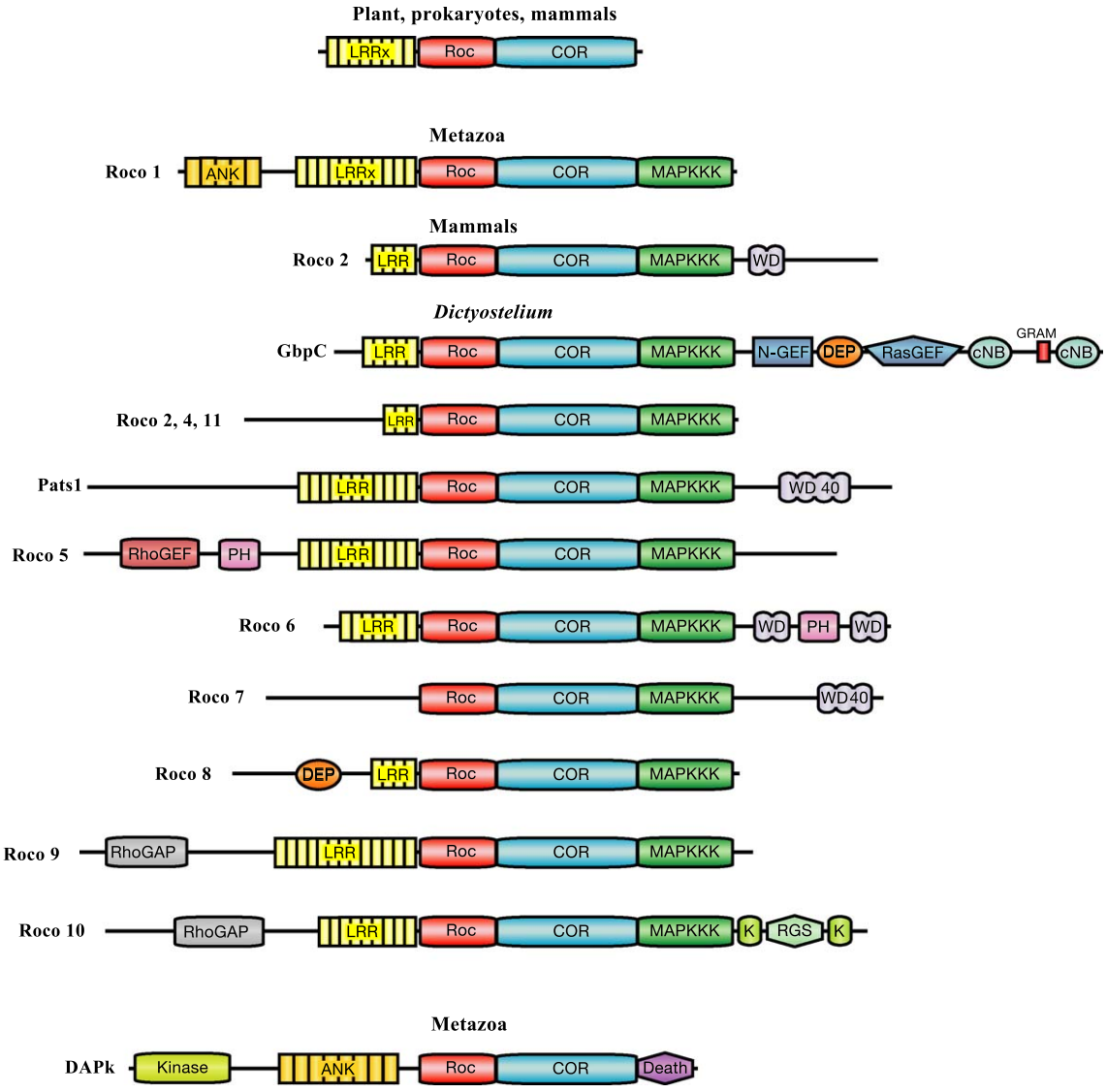

Fig. 1. Domain architecture of the Roco proteins. The domains are leucine-rich repeat (LRR), Ras in complex proteins domain (Roc), domain C-terminal of Roc (COR), ankyrin repeat (ANK), MAPkinase kinase kinase domain (MAPKKK), WD40 repeats (WD40), N-terminal motif of RasGEF (N-GEF), DEP domain, a domain with unknown function found in dishevelled, Egl-10 and Pleckstrin proteins (DEP), Ras guanine nucleotide exchange factor domain (RasGEF), GRAM domain, a domain in glucosyltransferases, myotubularins and other putative membrane-associated proteins (GRAM), cyclic nucleotide binding domain (cNB), Rho guanine nucleotide exchange factor domain (RhoGEF), Pleckstrin homology domain (PH), Rho GTPase activating protein domain (RhoGAP), Kelch motif (K), regulator of G protein signalling domain (RGS), Ser/Thr protein kinase domain (kinase), FAS/TNF cytosolic interaction domain (DEATH).

rich repeats (LRRs), which are 20-amino-acid-long repeats that are involved in protein–protein interactions [6].

The second group of Roco proteins is present in *Dicyostelium* and metazoa. In these proteins the COR domain is always succeeded by a kinase domain that belongs to the MAPKKK subfamily of kinases; the Roc domain is again preceded by 3–16 LRRs (see Fig. 1). In metazoan Roco1 proteins, the LRRs are preceded by five to seven Ankyrin (ANK) repeats, which are 33-amino-acid-long repeats that are also involved in protein–protein interactions [7]. In human, mouse and rat, a second Roco protein was identified that does not contain ANK repeats but has two WD40 repeats after the kinase domain. Surprisingly, in *Dictyostelium*, as many as nine new different homologous sequences were identified (GenBank AY232265, AY232266, AY232267, AY232268, AY232269, AY232270, AY232271, AY232272, AY232273), in addition to the previously mentioned GbpC, and Pats1, a recently described protein that plays a role in cytokinesis [8]. Apart from the LRRs, the Roc, COR and MAPKKK domains, *Dictyostelium* Roco proteins contain a remarkable diversity of other domains, including beta propeller forming repeats (WD40 and Kelch motif), and domains that interact with small G-proteins (RasGEF, RhoGEF and RhoGAP) or heterotrimeric G-proteins (RGS).

The third group of Roco proteins contains the exclusively metazoan tumour suppressor death-associated protein kinases (DAPk). The Roc domain of these well-characterised proteins has not been studied in detail, except for the recognition of its P-loop [9] or GTPase [10]. The COR domain of DAPk is followed by a death domain, which is found in proteins with apoptotic functions [11]. Furthermore, this is the only group of Roco proteins in which the Roc domain is not preceded by LRRs. Instead, seven to nine ANK repeats and a protein kinase domain are located at the N terminus. This kinase domain does not belong to the subgroup of MAPKKKs but is related to Ca^{2+}/calmodulin-regulated kinases.

3. Features of the Roc and COR domain

Although the Roc domain clearly belongs to the Ras/GTPase superfamily (see below), SMART and Pfam give poor expectation values for Ras/GTPase domains in the Roco proteins (expectation value above e − 06), and in fact Ras/GTPase domains are not recognized by SMART or Pfam in about half of the sequences. This is also the case for DAPk, which explains why the small GTPase domain has not been initially identified in this well studied protein.

The Roc and COR domains of the obtained sequences were aligned using the program ClustalW [12]. A full alignment including the MAPKKK domain can be viewed in Appendix A. The alignment of the Roc domains of 20 Roco proteins from pro- and eukaryotes was complemented with about nine members of each of the four well established groups of the Ras/GTPase superfamily, Ras, Rho/Rac, Arf and Rab/Ran [1]. Cluster analysis of this dataset (Fig. 2) reveals that Roc stands out as a separate monophyletic group of the Ras superfamily of small GTPases, clearly distinguished from the other four groups, which is supported by rather high bootstrap values. Within the Roc family, sequences from prokaryotes and eukaryotes are placed in separate groups, but bootstrap values at the base of the eukaryotes are rather low and not significant (see Fig. 2).

The small GTPases possess five loops with defined functions [13]. It appears that four loops are conserved in Roc/GTPases, including the amino acids involved in GTP-binding and hydrolysis (Fig. 3). The only exception is the G5 loop, which is mainly involved in the stabilisation of amino acids from the G4 loop that interact with the guanine nucleotide [14]. Accordingly, the amino acids in the G4 loop are also somewhat different in most of the Roc domains. Most striking is a histidine in all Roc domains from eukaryotes that is a lysine in conventional small GTPases. The methylene group of this lysine provides hydrophobic surface that lies over the purine ring, while in some structures the ϵ-amino group is hydrogen-bonded with an exocyclic oxygen of the ribose ring [15]. The

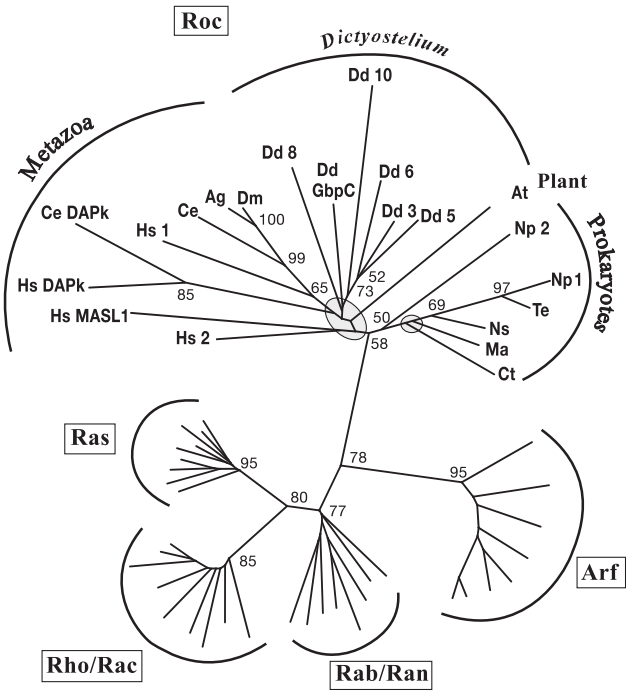

Fig. 2. Dendrogram of the Ras/GTPase superfamily containing 21 Roc domains and 34 other small G-protein domains from the four subgroups, Ras, Rab/Ran, Arf and Rho/Rac. The dendrogram was constructed with the Fitch program from the PHYLIP package [16] using an alignment that was created with ClustalW [12]. All 21 Roc domains were clustered in a monophyletic group; for clarity some were deleted from the figure. Numbers refer to bootstrap values and indicate that the distinction between the five groups is very reliable; within the Roc domain only bootstrap values above 50% are shown. Nodes with bootstrap values below 50% are present in two small regions of the tree, indicated by the grey areas. Species abbreviations are Hs, *Homo sapiens*; Ce, *Caenorhabditis elegans*; Dm, *Drosophila melanogaster*; Ag, *Anopheles gambiae*; Dd, *Dictyostelium discoideum*; At, *Arabidopsis thaliana*; Np, *Nostoc punctiforme*; Ct, *Chlorobium tepidum*; Te, *Trichodesmium erythraeum*; Ns, *Nostoc sp* (PCC 7120); Ma, *Methanosarcina acetivorans*.

histidine at this position may indicate an altered interaction between Roc and GTP. Roc domains have a few other common characteristics that separate them from other GTPases, such as an insertion between the P-loop and switch I (see Appendix A). Furthermore, they have a 17-amino-acid N-terminal extension (see Fig. 3) that is strongly predicted to form an α-helix according to the secondary structure prediction program PROF (http://www.embl-heidelberg.de/predictprotein/predictprotein.html). This extension is immediately preceded by an LRR in all Roco proteins except for DAPk.

The COR domain that succeeds the Roc domain is a 300–400-amino-acid-long region that does not show significant sequence homology to any domain or protein described today. Regions with high homology (Fig. 3) are frequently interrupted by insertions, especially in the Roco proteins from *Dictyostelium*. In the latter, the insertions are mostly repeats of asparagines and other hydrophilic amino acids, which is a common phenomenon

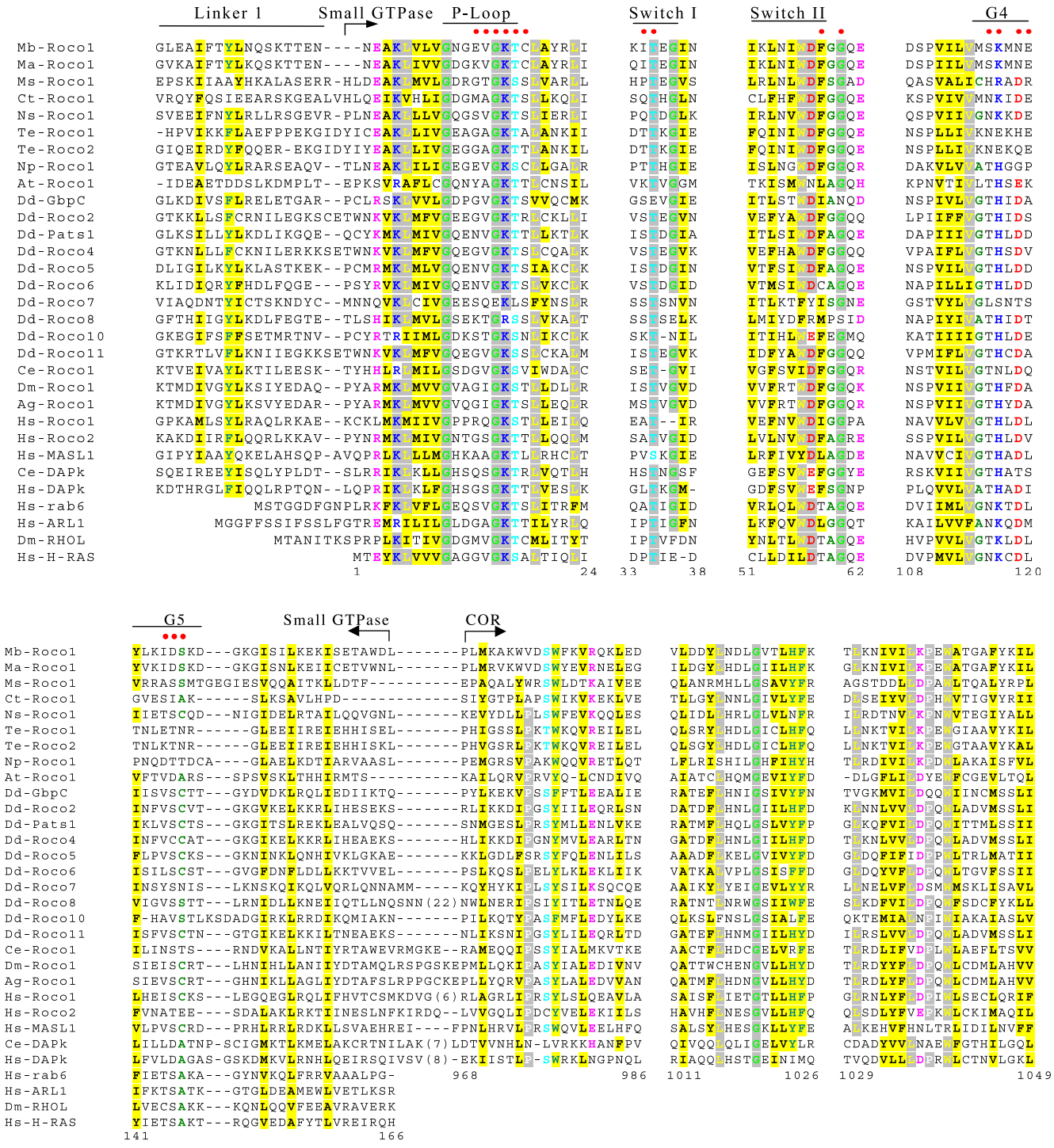

Fig. 3. Sequence alignment of some parts of the Roc and COR domains. The alignment shows the five loops that interact with GTP from 27 Roc domains and from one representative of the other four small GTPase subfamilies. The linker that connects the LRRs with the Roc domain is shown at the N terminus, and the connection with the COR domain is presented at its C terminus of Roc. Residues that interact with the nucleotide in the structure of Ha-Ras are marked with a red dot. Nine stretches of conserved sequence are presented for the COR domain, as well as its connection with the kinase domain. The amino acids of the Roc and COR domains are indicated by numbers, referring to the amino acid sequence of Hs-H-Ras and Hs-DAPk, respectively. The alignment was coloured according to a 75% consensus using CHROMA [17]: yellow background, hydrophobic (VLIMAFY); blue, positively charged (HKR); red, negatively charged (DE); magenta, charged (HKRDE); green, small (GASC); cyan, alcohol (ST); green on yellow, aromatic (FYWH); grey background indicates a conserved amino acid. See Fig. 2 for abbreviations.

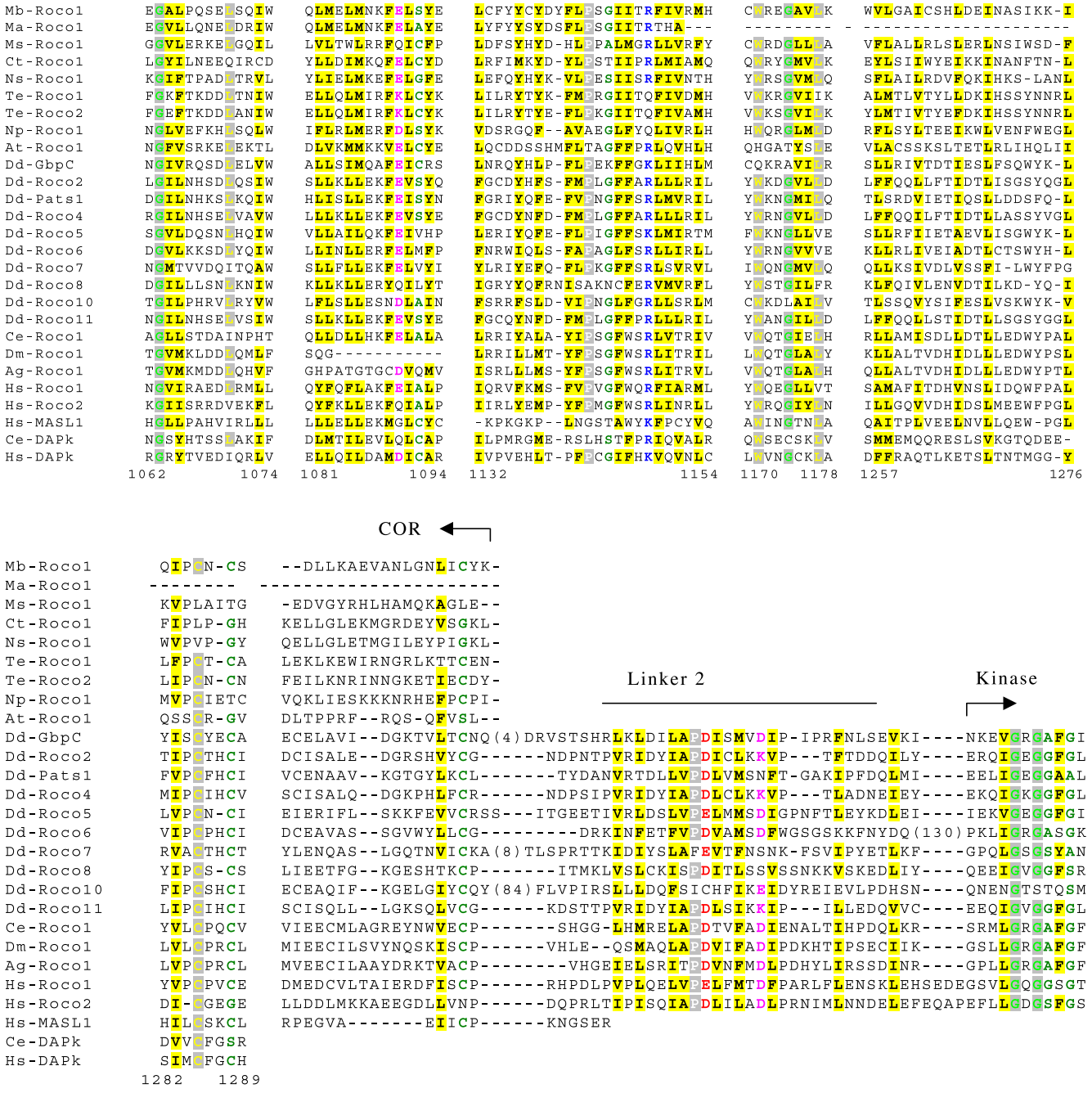

Fig. 3 (*continued*).

for proteins in this organism. The Roco proteins that contain a MAPKKK domain C-terminal of the COR domain all share a 30-amino-acid sequence motif just before the kinase domain that is absent in the other Roco proteins. Possibly, this region serves as a linker between the Roc/COR domains and the kinase domain or it may be part of the kinase itself.

The kinase domain that follows the COR domain in many Roco proteins is most closely related to the subfamily of MAPKKkinases; cluster analysis of about a hundred kinase domains strongly support the classification of the kinase domains of Roco proteins as MAPKKK (data not shown). Interestingly, one of the effectors of Ras proteins is MAPKKK, suggesting that the Roc domain might activate its own MAPKKK domain, thus acting as an intramolecular signalling cascade.

4. Function of Roco proteins

Three Roco proteins have been investigated to some extent until now, human DAPk and the *Dictyostelium*

proteins GbpC and Pats1. DAPk is a Ca^{2+}/calmodulin-regulated Ser/Thr kinase that positively influences apoptosis [9]. DAPk can phosphorylate myosin light chain, which results in membrane blebbing. Furthermore, it associates with actin microfilaments, which is essential for its death-promoting activity. Interestingly, the region that is essential and sufficient for this interaction (amino acid 641–835) overlaps with the first half of the Roc domain; it starts at 26 amino acids before the N-terminal extra α-helix that the Roc domains have, and ends shortly after the switch II region [9].

GbpC is a protein that is critical for cell polarity and chemotaxis [4]. It exerts its action, at least in part, by activating myosin II light chain phosphorylation, which is required for cell polarization. Since GbpC contains a kinase domain, we suggest that it can phosphorylate myosin kinases or an upstream kinase. Pats1 is a protein that was recently discovered that plays a role in cytokinesis [8]. Furthermore, it was observed that in cells lacking the *pats1* gene, myosin heavy chain did not properly localize to the cleavage furrow.

Although some members of the Ras/GTPase have been found in association with other domains, Roc is the first Ras-related subfamily that is exclusively present in complex proteins. Many intriguing questions are now to be addressed, such as how does its structure compare to that of the other members of the superfamily, how is the protein activated, what is the role of the COR domain, and especially what is the functional target of Roc. The MAPKKkinase domain provides an exciting candidate for intramolecular signal transduction. For the three Roco proteins studied so far, all seem to be involved in cytoskeletal rearrangements. However, since this group of proteins is rather diverse, it remains to be determined whether this is a general feature of the Roco proteins. With this in mind, it is remarkable that the Roc domain of DAPk is essential and sufficient for association of the protein to the cytoskeleton.

Appendix A

A full alignment of the Roc and COR domains can be found at http://www.rug.nl/scheikunde/onderzoek/programmas/cellBiochemistry/.

References

[1] Y. Takai, T. Sasaki, T. Matozaki, Small GTP-binding proteins, Physiol. Rev. 81 (2001) 153–208.
[2] D.D. Leipe, Y.I. Wolf, E.V. Koonin, L. Aravind, Classification and evolution of P-loop GTPases and related ATPases, J. Mol. Biol. 317 (2002) 41–72.
[3] J.M. Goldberg, L. Bosgraaf, P.J.M. Van Haastert, J.L. Smith, Identification of four candidate cGMP targets in *Dictyostelium*, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 6749–6754.
[4] L. Bosgraaf, H. Russcher, J.L. Smith, D. Wessels, D.R. Soll, P.J.M. Van Haastert, A novel cGMP signalling pathway mediating myosin phosphorylation and chemotaxis in *Dictyostelium*, EMBO J. 21 (2002) 4560–4570.
[5] T.L. Bailey, M. Gribskov, Combining evidence using *p*-values: application to sequence homology searches, Bioinformatics 14 (1998) 48–54.
[6] B. Kobe, A.V. Kajava, The leucine-rich repeat as a protein recognition motif, Curr. Opin. Struck. Biol. 11 (2001) 725–732.
[7] S.G. Sedgwick, S.J. Smerdon, The ankyrin repeat: a diversity of interactions on a common structural framework, Trends Biochem. Sci. 24 (1999) 311–316.
[8] J.C. Abysalh, L.L. Kuchnicki, D.A. Larochelle, The identification of Pats1, a novel gene locus required for cytokinesis in *Dictyostelium discoideum*, Mol. Biol. Cell 14 (2003) 14–25.
[9] O. Cohen, E. Feinstein, A. Kimchi, DAP-kinase is a Ca2+/calmodulin-dependent, cytoskeletal-associated protein kinase, with cell death-inducing functions that depend on its catalytic activity, EMBO J. 16 (1997) 998–1008.
[10] L. Aravind, V.M. Dixit, E.V. Koonin, Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons, Science 291 (2001) 1279–1284.
[11] N. Itoh, S. Nagata, A novel protein domain required for apoptosis. Mutational analysis of human Fas antigen, J. Biol. Chem. 268 (1993) 10932–10937.
[12] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[13] M. Paduch, F. Jelen, J. Otlewski, Structure of small G proteins and their regulators, Acta Biochim. Pol. 48 (2001) 829–850.
[14] H.R. Bourne, D.A. Sanders, F. McCormick, The GTPase superfamily: conserved structure and molecular mechanism, Nature 349 (1991) 117–127.
[15] S.R. Sprang, G protein mechanisms: insights from structural analysis, Ann. Rev. Biochem. 66 (1997) 639–678.
[16] J. Felsenstein, Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods, Methods Enzymol. 266 (1996) 418–427.
[17] L. Goodstadt, C.P. Ponting, CHROMA: consensus-based colouring of multiple alignments for publication, Bioinformatics 17 (2001) 845–846.